

Glauber Dynamics of colorings on trees

Allan Sly *

Yumeng Zhang †

Abstract

The mixing time of the Glauber dynamics for spin systems on trees is closely related to reconstruction problem. Martinelli, Sinclair and Weitz established this correspondence for a class of spin systems with soft constraints bounding the log-Sobolev constant by a comparison with the block dynamics [10, 11]. However, when there are hard constraints, the block dynamics may be reducible.

We introduce a variant of the block dynamics extending these results to a wide class of spin systems with hard constraints. This applies for essentially any spin system that has non-reconstruction provided that on average the root is not locally frozen in a large neighborhood. In particular we prove that the mixing time of the Glauber dynamics for colorings on the regular tree is $O(n \log n)$ in the entire known non-reconstruction regime.

1 Introduction

There has been substantial interest in understanding the rate of convergence of the Glauber dynamics for spin systems on trees and in particular how the mixing times scaling relates to the spatial mixing properties of the Gibbs measure. In the case of the coloring model, the natural conjecture is that there is rapid mixing (the mixing time is $O(n \log n)$) whenever the model is in the reconstruction regime. This was previously shown for the related block dynamics by Bhatnagar et al. [2] but the presence of frozen local regions means their proof does not apply to the single site dynamics. We overcome this restriction establishing the following result.

Theorem 1.1. *For fixed $\beta < 1 - \ln 2$ and $k > k(\beta)$ large enough when $d \leq k[\log k + \log \log k + \beta]$, the mixing time of Glauber dynamics of the k -coloring model on n -vertex d -ary tree is $O(n \log n)$.*

Our bound corresponds to the non-reconstruction region established in [15]. In the same paper it was shown that the k -coloring model is reconstructible for $\beta > 1$. In a forthcoming work we will give an improved upper bound on the reconstruction threshold from which it will follow that $O(n \log n)$ mixing holds for the full non-reconstruction regime.

*University of California, Berkeley and Australian National University. Supported by an Alfred Sloan Fellowship and NSF grant DMS-1208338. Email: sly@stat.berkeley.edu

†University of California, Berkeley. Email: zym3008@berkeley.edu

1.1 Previous work

Many studies have shown that the mixing time of Glauber dynamics, both for k -coloring model and general spin systems, are related to spatial properties of the Gibbs measure. In the case of d -ary trees, two properties of primary interest are uniqueness of infinite volume Gibbs measure and reconstruction, which corresponds to the extremality of infinite Gibbs measure induced by free boundary conditions. It has been shown in quite general settings that the Glauber dynamics exhibits rapid mixing, and in particular an $O(n \log n)$ mixing time, when the system admits a unique infinite volume Gibbs measure by Martinelli, Sinclair and Weitz [10, 11] and most generally in Weitz's thesis [18]. Beyond the uniqueness threshold, the Glauber dynamics for k -coloring may not even be ergodic under some boundary conditions, but it is natural to consider the behavior under free boundary conditions.

There have been intensive studies on the mixing time of coloring model from both theoretical computer science and statistical physics. For k -coloring on general graphs with n vertices and maximal degree d , the Glauber dynamics is irreducible when $k \geq d + 2$. It is tempting to conjecture that the chain exhibits rapid mixing whenever $k \geq d + 2$. So far the best result on general graphs is given by Vigoda in [17], where he showed $O(n^2 \log n)$ mixing time when $k \geq \frac{11}{6}d$. A series of improvements on the constant $\frac{11}{6}$ for rapid mixing have been made with extra conditions on the degree or girth, (see the survey [4] for more results toward this direction).

For the k -coloring model on d -ary trees, Jonasson [7] established there is a unique infinite Gibbs measure for $k \geq d + 2$. Martinelli et al. [10, 11, 18] studied the mixing rates of several spin systems on d -ary trees and in particular showed $O(n \log n)$ mixing time for coloring models under any boundary conditions for $k \geq d + 3$. Their method made use of the block dynamics and employed the idea of the decay of correlation between the root and the leaves. As k drops below the uniqueness threshold of $d + 2$, we can no longer compare the block dynamics to the standard single site Glauber dynamics as the chain inside a block will not always be connected under worst-case boundary conditions. Notwithstanding this, the desired speed of decay of correlation still holds suggesting that rapid mixing under free boundary conditions might still be true beyond the non-uniqueness threshold.

This decay of correlation between root and leaves is closely related to another problem called reconstruction. Roughly speaking, a model on tree is reconstructible if, given the leaves of a randomly chosen configuration, one's best guess for the root is "strictly better" than the stationary distribution, as the number of level goes to infinity. In other words, reconstruction corresponds to the non-vanishing influence of average-case boundary conditions to the root. For coloring models, Mossel and Peres [13] established reconstruction when $k \geq (1 + o(1))d / \log d$ by considering the point at which the model freezes, that is the boundary condition exactly determines the root. In the other direction, Bhatnagar et al. [2] and Sly [15] independently proved that the model has non-reconstructible for $k \leq (1 + o(1))d / \log d$. Using this result [2] showed that the block dynamics for k -coloring model mixes in $O(n \log n)$ time for $k \geq (1 + \epsilon)d / \log d$ and large $k > k(\epsilon)$ using non-reconstruction and following the methods of [10].

For more results below the non-uniqueness threshold, Berger et al. [1] showed polynomial mixing time for general models on trees whenever the dynamics is ergodic, which in the case of coloring corresponds to $k \geq 3$

and $d \geq 2$. For the coloring model Goldberg et al. [5] proved an upper bound of $n^{O(d/\log d)}$ for the complete tree for with branching factor d and Lucier et al. [8] showed the mixing time is $n^{O(1+d/k \log d)}$ for all d and $k \geq 3$. Recently Tetali et al. [16] proved the mixing time undergoes a phase transition at the reconstruction threshold $k = (1 + o(1))d/\log d$, where their upper bound for $k \geq (1 + o(1))d/\log d$ is $O(n^{1+o_k(1)})$. They also showed that the mixing time is $\Omega(n^{d/k \log d - o_k(1)})$ for $k \leq (1 - o(1))d/\log d$, which implies rapid mixing does not hold in reconstruction region.

The main result of this paper is to reduce the mixing time in the non-reconstruction region from the polynomial time bound of $n^{1+o(1)}$ to a sharp bound of $O(n \log n)$. Our proof is a modification of the techniques used in [10]. The main technical difficulty of directly applying their method is their (non-obvious) restriction of “admissible” and “well-connected”, i.e. for any boundary conditions on arbitrary subset the set of proper configurations is non-empty and connected via single site update. These conditions may not be satisfied in models with hard constraints, in particular k -coloring model. Intuitively, below the uniqueness threshold, there will be vertices whose states are “frozen” by their neighbors. While block dynamics can update “frozen” vertices together with their neighbors in a single move, extra efforts are needed for single site dynamics to pass around the barrier and change them, leading to the failure of comparing two dynamics. We will look at a new variant of block dynamics that focuses on the connected component induced by the single site chain on the state space of usual block dynamics. By carefully examining the portion of “frozen” vertices and their influence on nearby sites, we will show rapid mixing of our new version of block dynamics which in turn implies the final result.

1.2 General spin system

Phase transitions and decay of correlation of spin systems play a key role in the mixing time of the Glauber dynamics. On trees the reconstruction threshold rather than the uniqueness threshold plays the key role. Berger et al. [1] showed that for general spin systems, $O(n)$ relaxation time under free-boundary condition implies non-reconstruction. In the other direction, Weitz conjectured [18] that for any k -state spin system or on d -ary trees, the system mixes in $O(n \log n)$ time whenever it admits a unique Gibbs measure and the Glauber dynamics is connected under given boundary condition. He proved for the case $k = 2$ and for ferromagnetic Potts model and coloring as two special cases of $k > 2$. His proof also gives spatial mixing conditions which apply for a wide range of models.

Our result for k -coloring can be extended to general k -state spin systems and gives a sufficient condition, for spin systems to exhibit rapid mixing in the non-reconstruction region. Throughout the paper, we will refer to a spin system by its probability kernel M , defined by $M(c, c') = \mu(\sigma_y = c' | \sigma_x = c), (x, y) \in E$. The sufficient condition, which we will call the **connectivity condition** \mathcal{C} is specified in Section 2.2 and deals with hard constraints. It roughly speaking asks that the probability that the root can “change freely” to all k states given a random boundary condition tends to 1. In particular, it is automatically satisfied by all models without hard constraints or with a permissive state – one that can occur next to any other (e.g. the hardcore model).

Theorem 1.2. *Let M be a k -state system on the n -vertex d -ary tree T with second eigenvalue λ . If M satisfies the connectivity condition \mathcal{C} , is non-reconstructible on T , and $d\lambda^2 < 1$ then the mixing time of Glauber dynamics on T under free boundary condition is $O(n \log n)$.*

The assumption that $d\lambda^2 < 1$ corresponds to the Kesten-Stigum bound in the reconstruction problem. When $d\lambda^2 > 1$, the system is always reconstructible by counting the number of leaves with each state (see, the survey [12]). Hence non-reconstruction implies $d\lambda^2 \leq 1$. Our assumption that $d\lambda^2$ is strictly less than 1 is essential to our proof and indeed must be so since Ding, Lubetzky and Peres [3] showed that the mixing time for the Ising model is at least of order $n \log^3 n$ when $d\lambda^2 = 1$.

2 Preliminaries

2.1 Definition of model

Spin systems: We will use $[k] = \{1, \dots, k\}$ to denote the state space of a spin and $T = (V, E)$ to denote the d -ary tree (i.e. every vertex have d offsprings) with root ρ and $|T| = n$ vertices. Throughout the paper, we will denote the l -th level of the tree T by L_l , with $L_0 = \{\rho\}$. We will also use T_x to represent the subtree rooted at $x \in T$ and let $B_{x,l}$, $L_{x,l}$ denote the first l levels and the l -th level of T_x respectively.

A configuration on T is an assignment of spins to vertices $\sigma \in [k]^V$. For a k -state spin system (with potential U and W), the probability of seeing $\sigma \in [k]^V$ is given by the (free-boundary) Gibbs measure

$$\mu(\sigma) = \frac{1}{Z} \exp \left[- \left(\sum_{(x,y) \in E} U(\sigma_x, \sigma_y) + \sum_{x \in V} W(\sigma_x) \right) \right],$$

where U is a symmetric function from $[k] \times [k] \rightarrow \mathbb{R} \cup \{\infty\}$, W is a function from $[k] \rightarrow \mathbb{R} \cup \{\infty\}$ and Z is the partition function independent of σ such that $\sum_{\sigma \in [k]^V} \mu(\sigma) = 1$. If for some states $i, j \in [k]$, $U(i, j) = \infty$, we say (i, j) is a hard constraint, otherwise we say i and j are *compatible*. We will focus on the set of proper configurations on T , denoted by $\Omega_T = \{\sigma, \mu(\sigma) > 0\}$. For example in the coloring model, $W(c) = 0$, $U(c, c') = \infty \cdot 1(c = c')$ and the Gibbs measure is the uniform distribution over all proper colorings. We will write σ_A for the restriction of σ to the subset A and use superscript to denote conditioning on a boundary condition. $\Omega_A^\eta = \{\sigma, \sigma \in \Omega_T, \sigma_{T \setminus A} = \eta_{T \setminus A}\}$ is the set of configurations compatible with boundary condition η and the conditional law is $\mu_A^\eta(\sigma) = \mu(\sigma \mid \sigma \in \Omega_A^\eta)$.

The principal example for this paper is the proper graph coloring. A proper k -coloring of graph $G = (V, E)$ is an assignment $\sigma : V \rightarrow [k] = \{1, 2, \dots, k\}$ such that for all $(x, y) \in E$, $\sigma_x \neq \sigma_y$. In the statistical physics literature this corresponds to zero-temperature anti-ferromagnetic Potts model.

For the reconstruction problem, it is easy to work with the Markov chain construction of Gibbs measure on trees, which can be taken as a special case of the broadcast model on trees. We think of the process where

information is sent on tree T from the root ρ downwards and each edge acts as a noisy channel. For each input $c_1 \in [k]$, the output of c_2 is chosen from probability kernel $M(c_1, c_2)$. If the input at root ρ follows the stationary distribution of M , denoted by π , the law of a random configuration on T is given by

$$\mu(\sigma) = \pi_{\sigma_\rho} \prod_{(x,y) \in E} M(\sigma_x, \sigma_y).$$

It is easy to check the following one-to-one correspondence between potential U, W and reversible probability kernel M ,

$$M(c_1, c_2) = \frac{\exp[-(U(c_1, c_2) + W(c_2))]}{\sum_{c' \in [k]} \exp[-(U(c_1, c') + W(c'))]}, U(c_1, c_2) = \ln \left(\frac{M(c_1, c_2)}{\pi_{c_2}} \right), W(c) = -\ln \pi_c.$$

We will henceforth denote a spin system by its probability kernel M , we only deal with M that are ergodic and reversible.

2.1.1 Glauber dynamics and mixing time

The Glauber dynamics for a k -state spin system M is a Markov chain X_t on state space Ω_T . A step of the Markov chain from X_t to X_{t+1} is defined as follows:

1. Pick a vertex x uniformly randomly from T ;
2. Pick a state $c \in [k]$ according to the conditional distribution of the spin value of x given the rest of configuration, i.e. state c is picked with probability $\mu_{\{x\}}^\sigma(c) = \mu(\sigma'_x = c \mid \sigma'_y = \sigma_y, \forall y \neq x)$;
3. Set $X_{t+1}(x) = c$ and $X_{t+1}(y) = X_t(y), \forall y \neq x$.

We denote the transition matrix by P . In the case of coloring, the second step corresponds to a picking a uniformly chosen color that does not appear in the neighbor of x .

As we will show in Lemma 2.1, under the connectivity condition \mathcal{C} , the Glauber dynamics with free boundary conditions is ergodic, reversible and hence converges to the Gibbs measure μ (this is easy to check for models with no hard constraints). The **mixing time** is defined as

$$t_{\text{mix}} = \max_{\sigma \in \Omega_T} \min\{t : d_{\text{TV}}(P^t(\sigma, \cdot), \mu) \leq \frac{1}{2e}\},$$

where P is the probability kernel of X_t and $d_{\text{TV}}(\pi, \mu) = \frac{1}{2} \sum_{\sigma} |\pi(\sigma) - \mu(\sigma)|$ is the total variance distance. To bound the mixing time we will make use of the **log-Sobolev constant**. For non-negative function $f : \Omega_T \rightarrow \mathbb{R}$, let $\mu(f) = \sum_{\sigma} \mu(\sigma) f(\sigma)$ be the expectation of f and the $\text{Ent} f = \mu(f \log f) - \mu(f) \log \mu(f)$ be the entropy. The

Dirichlet form of f is defined by

$$\mathcal{D}(f) = \frac{1}{2} \sum_{\sigma, \sigma'} \mu(\sigma) P(\sigma, \sigma') (f(\sigma) - f(\sigma'))^2.$$

And the log-Sobolev constant is defined by $\gamma = \inf_{f \geq 0} \frac{\mathcal{D}(\sqrt{f})}{\text{Ent}(f)}$. Applying results in functional analysis to Glauber dynamics yields the following bound on mixing time regarding log-Sobolev constant (see e.g. Theorem 2.2.5 of [14]):

Theorem. *For k -state system M on n -vertex d -ary tree T , there exists a constant $C > 0$ such that $t_{\text{mix}} \leq \frac{1}{\gamma} \cdot Cn \log n$.*

Therefore to show rapid mixing is enough to show that γ is uniformly bounded away from zero as n tends to infinity.

2.1.2 Uniqueness and reconstruction

Two key notions of spatial decay of correlation for spin systems on trees are the uniqueness and reconstruction thresholds. Recalling that L_l is the set of the vertices at level l in T , we have the following definition

Definition (Reconstruction). A k -state system M is reconstructible on a tree T if for some $c, c' \in [k]$

$$\limsup_{l \rightarrow \infty} d_{TV}(\mu(\sigma_{L_l} = \cdot \mid \sigma_\rho = c), \mu(\sigma_{L_l} = \cdot \mid \sigma_\rho = c')) > 0.$$

Otherwise we say the system has non-reconstruction on T .

Non-reconstruction is equivalent to the externality of infinite volume Gibbs measure under free boundary conditions. More equivalent definition and an extensive literature are given in the survey [12]. A strictly stronger condition is the uniqueness property.

Definition (Uniqueness). We say a k -state system M has uniqueness on a tree T if

$$\limsup_{l \rightarrow \infty} \sup_{\eta, \eta' \in \Omega_{L_l}} d_{TV}(\mu(\sigma_\rho = \cdot \mid \sigma_{L_l} = \eta), \mu(\sigma_\rho = \cdot \mid \sigma_{L_l} = \eta')) > 0,$$

where Ω_{L_l} is the set of configurations on level l .

2.2 Connectivity condition

In this section we specify the connectivity condition \mathcal{C} . First we will define the notion for a vertex to be free. Let T be a tree of l levels. For configuration $\sigma \in \Omega_T$ with $\sigma_\rho = c$, $\sigma_{L_l} = \eta$, we say the root can change (from c)

to state c' **in one step** if and only if there exist a path $\sigma = \sigma^0, \sigma^1, \dots, \sigma^n \in \Omega_T$ such that for each i , σ^i, σ^{i+1} differ at exactly one vertex, $\sigma_{L_l}^i \equiv \eta$ for $\sigma_\rho^i = c, i = 0, 1, \dots, n-1, \sigma_\rho^n = c'$. Put another way, the path is a valid trajectory of the Glauber dynamics with fixed leaves which changes the state of ρ only once in the final step. For $x \in T$, we say x is **free** (in σ) if, considered as root of T_x , x can be changed to all the other $(k-1)$ -states in one step. Denote the probability that the root of an l -level tree is free as $p_l^{\text{free}} = \mu(\sigma : \rho \text{ is free in } \sigma)$.

Definition. We say that the k -state system M on the d -ary tree satisfies the connectivity condition \mathcal{C} if M is ergodic, reversible and

1. For all $c_1, c_2, c_3 \in [k]$, there exists $c \in [k]$ such that c is compatible with c_1, c_2, c_3 .
2. The probability of being free tends to 1 as l tends to infinity, i.e. $\lim_{l \rightarrow \infty} p_l^{\text{free}} = 1$.

The first condition is used in the proof of Claim 4.2, see the discussion afterward for the necessity of this condition.

We first show that under connectivity condition \mathcal{C} , the Glauber dynamic is irreducible under free boundary conditions and ergodicity follows from that. For the sake of recursive analysis on subtrees later, we want to consider the case where the parent of the root is fixed to be some state. For state $c \in [k]$, let Ω_T^c denote the set of configurations with the parent of root ρ being state c and let μ_T^c be the corresponding conditional Gibbs measure.

Lemma 2.1. *For general k -state system M on d -ary tree T . If M is ergodic and reversible, then Ω_T^c is irreducible under Glauber dynamics.*

Proof. Reversibility follows from the detailed balance equations. We prove irreducibility by induction on the number of levels l in T . For $l = 0$, it is trivially true. We assume that the Glauber dynamics is connected for $(l-1)$ -level tree and consider l -level tree T and configurations $\sigma, \sigma' \in \Omega_T^c$. To establish a path of valid moves of the dynamics from σ to σ' , one can first change every vertex $x \in L_1$ to state c by a sequence of moves in the tree T_x using our inductive assumption. We may then change the spin of the root from σ_ρ to σ'_ρ . Finally we may change the configuration of every subtree T_x to σ'_{T_x} using the inductive assumption bringing us to the configuration σ' . \square

2.3 Component Dynamics

In this section, we define a new variant of block dynamics on T , namely “the component dynamics”, which updates a block of vertices each step, but only chooses configurations within the connected component of the Glauber dynamic. In this way we can utilize the techniques in [10] while bypassing the problem that steps of block dynamic may not be connected in Glauber dynamics when $k \leq d+1$. To give a formal definition, for $A \subset T$, we say $\sigma' \sim_A \sigma$ if $\sigma'_{T \setminus A} = \sigma_{T \setminus A}$ and σ'_A, σ_A are connected in Glauber dynamics on A with fixed boundary condition $\sigma_{T \setminus A}$. We will omit the A in $\sigma \sim_A \sigma'$ when it is clear from the context. Let $\Omega_A^{*,\sigma} = \{\sigma' \in \Omega_A^\sigma, \sigma' \sim_A \sigma\}$

denote the connected component of σ in Ω_A^σ , and let $\mu_A^{*,\sigma}(\sigma') = \mu(\sigma' | \Omega_A^{*,\sigma})$ be the Gibbs distribution conditioned on both configuration outside A and the connected component within A .

For $l \geq 1$, recall $B_{x,l}$ is the block of l levels rooted at x and $L_{x,l}$ be the l -th level of $B_{x,l}$. If x is within distance l of the leaves, let $B_{x,l} = T_x$. We define a step of the **component dynamics** by the update rule:

1. Pick a vertex x uniformly randomly from T ,
2. Replace σ by σ' drawn from conditional distribution $\mu_{B_{x,l}}^{*,\sigma}$.

The dynamics is reversible with respect to the Gibbs distribution. For test function $f : \Omega_T \rightarrow \mathbb{R}$, let $\mu_A^{*,\sigma}(f) = \sum_{\sigma' \in \Omega_A^{*,\sigma}} f(\sigma') \mu_A^{*,\sigma}(\sigma')$ be the conditional expectation of f on $\Omega_A^{*,\sigma}$ and for $f \geq 0$, let

$$\text{Ent}_A^{*,\sigma}(f) = \text{Ent}(f | \Omega_A^{*,\sigma}) = \mu_A^{*,\sigma}(f \log f) - \mu_A^{*,\sigma}(f) \log \mu_A^{*,\sigma}(f)$$

be the conditional entropy of f . We write the sum of local entropies of block size l as $\mathcal{E}_l^* = \sum_{x \in T} \mu_T(\text{Ent}_{B_{x,l}}^{*,\sigma}(f))$. With minor modification, the comparison result of block dynamics also works for component dynamics: (see e.g. Prop 3.4 of [9], in the proof substitute $\mathcal{E}_D(f, f)$ by \mathcal{E}_l^* and note $\sum_{\sigma'} \mu_T^{\sigma'}(\sigma') \mu_{B_{x,l}}^{*,\sigma'}(\sigma) = \mu_T^{\sigma'}(\sigma)$.)

$$\gamma \geq \frac{1}{l} \cdot \inf_{f \geq 0} \frac{\mathcal{E}_l^*}{\text{Ent}(f)} \cdot \min_{\sigma, x} \gamma_{B_{x,l}}^{*,\sigma}$$

where $\gamma_{B_{x,l}}^{*,\sigma}$ is the log-Sobolev constant of Glauber dynamics on $\Omega_{B_{x,l}}^{*,\sigma}$ with boundary condition on $\partial B_{x,l}$ given by σ . From our definition of $\Omega_{B_{x,l}}^{*,\sigma}$, it is easy to see that $\min_{\sigma, x} \gamma_{B_{x,l}}^{*,\sigma}$ is a constant only depending on the branching number d , block size l and M itself and is strictly greater than 0 independent of T . Thus to show $O(n \log n)$ mixing time for Glauber dynamics, it is enough to show $\mathcal{E}_l^* \geq \text{const} \times \text{Ent}(f)$ for all $f \geq 0$ and some choice of block size l independent of tree size $|T| = n$.

2.4 Outline of Proof

A key ingredient in [10] is that a certain strong concentration property implies “entropy mixing” in space which in turn implies the fast mixing of block dynamics. The following Theorem 2.2 can be seen as the combination of Theorems 3.4 and 5.3 of [10] adapted to component dynamics (the notation here is closer to Theorem 5.1 of [2]). For completeness, we include an outline of the proof in Section 5 pointing out the differences from the original argument.

Theorem 2.2. *There exist some constant $\alpha > 0$ such that for every $\delta > 0$ and $l \geq 1$, if for all $x \in T$ that is at least l levels from the leaves and any compatible pair of states $c, c' \in [k]$, $\mu^c = \mu_{T_x}^c$ satisfies*

$$\Pr_{\tau \sim \mu^c} \left(\left| \frac{\mu^c(\sigma_x = c' | \sigma \sim_{B_{x,l}} \tau)}{\mu^c(\sigma_x = c')} - 1 \right| \geq \frac{(1-\delta)^2}{\alpha(l+1-\delta)^2} \right) \leq e^{-2\alpha(l+1-\delta)^2/(1-\delta)^2} \quad (2.1)$$

then for every function $f \geq 0$, $\text{Ent}(f) \leq \frac{2}{\delta} \mathcal{E}_l^*$.

To prove Theorem 1.2, it suffices to verify (2.1) for some choice of l and δ . Note the original inequality in Theorem 5.3 of [10] or Theorem 5.1 of [2] is

$$\Pr_{\tau \sim \mu^c} \left(\left| \frac{\mu^c(\sigma_x = c' \mid \sigma_{L_{x,l}} = \tau_{L_{x,l}})}{\mu^c(\sigma_x = c')} - 1 \right| \geq \frac{(1-\delta)^2}{\alpha(l+1-\delta)^2} \right) \leq e^{-2\alpha(l+1-\delta)^2/(1-\delta)^2} \quad (2.2)$$

The only difference between (2.1) and (2.2) is that in equation (2.1), the inner measure μ^c conditions not only on boundary condition $\sigma_{L_{x,l}} = \tau_{L_{x,l}}$, but also the connected component of τ . We will first prove a stronger version the original inequality given the non-reconstruction of system M and $d\lambda^2 < 1$.

Theorem 2.3. *For a k -state system M , if M is non-reconstructible and $d\lambda^2 < 1$, then there exist some constant $\xi > 0$, $l_0 \geq 1$ such that for all $l \geq l_0$, every $x \in T$ that is at least l levels from the leaves, and any compatible pair of states $c, c' \in [k]$, $\mu^c = \mu_{T_x}^c$ satisfies*

$$\Pr_{\tau \sim \mu^c} \left(\left| \frac{\mu^c(\sigma_x = c' \mid \sigma_{L_{x,l}} = \tau_{L_{x,l}})}{\mu^c(\sigma_x = c')} - 1 \right| \geq e^{-\xi l} \right) \leq \exp(-e^{\xi l}). \quad (2.3)$$

From there, we will show under our connectivity condition, the difference between $\sigma \sim_{B_{x,l}} \tau$ and $\sigma_{L_{x,l}} = \tau_{L_{x,l}}$ is negligible in the upper half of the block when l is large and hence a similar tail distribution holds for the root.

Lemma 2.4. *Let M be a k -state system satisfying \mathcal{C} such that (2.2) holds for $l \geq l_0$ and $\delta = \delta_0$. Then there exist constants $l_1 \geq 2l_0$ and $\delta_1 \geq \delta_0$ such that for all $l \geq l_1$, equation (2.1) holds with $\delta = \delta_1$.*

Theorems 2.2 and 2.3 and Lemma 2.4 together imply Theorem 1.2. The rest of the paper is structured as follows, we will prove Theorem 2.3 in Section 3 and Lemma 2.4 in Section 4, and we will include a sketch of Theorem 2.2 in Section 5. After that we will apply the result to the k -coloring model and prove Theorem 1.1 in Section 6.

3 From non-reconstruction to the strong concentration property

In this section we prove Theorem 2.3. The result for k -coloring model was proved in [2], which used the specific structure of coloring model. Here we will give a different proof for general systems M using only non-reconstruction and that M is strictly below Kesten-Stigum bound $d\lambda^2 < 1$. We first introduce some notations. Recall that the stationary distribution of M is π . For $x \in T$, let

$$\tilde{R}_{x,l}(\tau)(c) = \frac{1}{\pi_c} \mu_{T_x}(\sigma_x = c \mid \sigma_{L_{x,l}} = \tau_{L_{x,l}})$$

denote the ratio of conditional and unconditional distribution at x and write $R_{x,l}(\tau) = \|\tilde{R}_{x,l}(\tau) - 1\|_\infty = \max_{c \in [k]} |\tilde{R}_{x,l}(\tau)(c) - 1|$. We will omit τ when it is clear from context. In the proof we will work with the

unconditional Gibbs measure $\mu = \mu_{T_x}$ and π instead of $\mu_{T_x}^c$ and $\mu_{T_x}^c(\sigma_x = c')$ and show the following inequality

$$\Pr_{\tau \sim \mu}(R_{x,l}(\tau) \geq e^{-\xi^l}) \leq \exp(-e^{\xi^l}). \quad (3.1)$$

To see (3.1) implies (2.3), consider the Markov chain construction of σ . Let E be the edge set of T_x , we have

$$\mu(\sigma) = \pi_{\sigma_x} \prod_{(y,z) \in E} M(\sigma_y, \sigma_z), \quad \mu^c(\sigma) = M(c, \sigma_x) \prod_{(y,z) \in E} M(\sigma_y, \sigma_z).$$

Hence the Radon-Nikodym derivative $\frac{d\mu^c}{d\mu}$ satisfies $\frac{d\mu^c}{d\mu}(\sigma) = \frac{M(c, \sigma_x)}{\pi_{\sigma_x}} \leq \pi_{\min}^{-1}$, where $\pi_{\min} = \min_{c \in [k]} \pi_c$. Note

$$\left| \frac{\mu^c(\sigma_x = c' \mid \sigma_{L_{x,l}} = \tau_{L_{x,l}})}{\mu^c(\sigma_x = c')} - 1 \right| = \left| \frac{1}{\pi_{c'}} \mu(\sigma_x = c' \mid \sigma_{L_{x,l}} = \tau_{L_{x,l}}) - 1 \right| \leq R_{x,l}(\tau),$$

we have that the LHS of equation (2.3) can be upper bounded by the LHS of (3.1) up to a factor of π_{\min}^{-1} . The following lemma gives the recursive relation of $\tilde{R}_{x,l}(c)$.

Lemma 3.1. *Let x_1, \dots, x_d denote the d children of $x \in T$, $\tilde{R}_{x,l}$ can be written as a function of $\tilde{R}_{x_i, l-1}$:*

$$\tilde{R}_{x,l}(c) = \frac{\prod_{i=1}^d M \tilde{R}_{x_i, l-1}}{\pi \prod_{i=1}^d M \tilde{R}_{x_i, l-1}}(c) = \frac{\prod_{i=1}^d \sum_{c_i \in [k]} M(c, c_i) \tilde{R}_{x_i, l-1}(c_i)}{\sum_{c' \in [k]} \pi_{c'} \prod_{i=1}^d \sum_{c_i \in [k]} M(c', c_i) \tilde{R}_{x_i, l-1}(c_i)}. \quad (3.2)$$

Proof. Let E and E_i denote the edge set of T_x and T_{x_i} , they satisfy $E = \cup_i (E_i \cup \{(x, x_i)\})$. Let $\Omega(c) = \{\sigma : \sigma_x = c, \sigma_{L_{x,l}} = \tau_{L_{x,l}}\}$ and $\Omega_i(c_i) = \{\sigma : \sigma_{x_i} = c_i, \sigma_{L_{x_i, l-1}} = \tau_{L_{x_i, l-1}}\}$ be the set of configurations on T_x and T_{x_i} with boundary condition τ . By the Markov chain construction, we have

$$\begin{aligned} \mu(\Omega(c)) &= \mu(\sigma_x = c, \sigma_{L_{x,l}} = \tau_{L_{x,l}}) \\ &= \sum_{\sigma \in \Omega(c)} \pi_c \prod_{(y,z) \in E} M(\sigma_y, \sigma_z) = \sum_{c_1, \dots, c_d \in [k]} \pi_c \prod_{i=1}^d M(c, c_i) \sum_{\sigma^i \in \Omega_i(c_i)} \prod_{(y,z) \in E_i} M(\sigma_y^i, \sigma_z^i) \\ &= \sum_{c_1, \dots, c_d \in [k]} \pi_c \prod_{i=1}^d \frac{M(c, c_i)}{\pi_{c_i}} \mu(\Omega_i(c_i)) = \pi_c \prod_{i=1}^d \sum_{c_i \in [k]} \frac{M(c, c_i)}{\pi_{c_i}} \mu(\Omega_i(c_i)). \end{aligned}$$

Therefore by Bayes formula,

$$\begin{aligned} \tilde{R}_{x,l}(c) &= \frac{1}{\pi_c} \mu(\sigma_x = c \mid \sigma_{L_{x,l}} = \tau_{L_{x,l}}) = \frac{1}{\pi_c} \frac{\mu(\Omega(c))}{\sum_{c' \in [k]} \mu(\Omega(c'))} \\ &= \frac{\prod_{i=1}^d \sum_{c_i \in [k]} \frac{M(c, c_i)}{\pi_{c_i}} \mu(\Omega_i(c_i))}{\sum_{c' \in [k]} \pi_{c'} \prod_{i=1}^d \sum_{c_i \in [k]} \frac{M(c', c_i)}{\pi_{c_i}} \mu(\Omega_i(c_i))} = \frac{\prod_{i=1}^d \sum_{c_i \in [k]} M(c, c_i) \tilde{R}_{x_i, l-1}(c_i)}{\sum_{c' \in [k]} \pi_{c'} \prod_{i=1}^d \sum_{c_i \in [k]} M(c', c_i) \tilde{R}_{x_i, l-1}(c_i)}. \end{aligned}$$

where the last step followed by dividing both the nominator and denominator by $\prod_{i=1}^d \sum_{c'_i \in [k]} \mu(\Omega_i(c'_i))$. \square

Observe that in the recursive relationship of (3.2), $\tilde{R}_{x,l}(c)$ is a rational function of $(\tilde{R}_{x_i,l-1}(c_i))_{i=1,\dots,d,c_i \in [k]}$. If for all $i = 1, \dots, d$ we have $\tilde{R}_{x_i,l-1} = 1$, then $\tilde{R}_{x,l} = 1$. By the continuity of (3.2) and ergodicity of M , we establish the following contraction property of $R_{x,l}$.

Lemma 3.2. *There exist an integer $m \geq 1$ and constant $\epsilon > 0$ such that for all d^m vertices $y_1, \dots, y_{d^m} \in L_{x,m}$, if at most one y_i has $R_{y_i,l-m} > \epsilon$ then*

$$R_{x,l} \leq \frac{1}{2} \sum_{i=1}^{d^m} R_{y_i,l-m}. \quad (3.3)$$

Proof. Denote the range of $\tilde{R}_{x,l}$ by the k dimensional simplex $\Delta_{[k]} = \{R \in \mathbb{R}^k, \pi R = 1, R_i \geq 0, i = 1, \dots, k\}$. Let $f : \Delta_{[k]}^d \rightarrow \Delta_{[k]}$ be the function on the RHS of (3.2) such that $\tilde{R}_{x,l} = f(\tilde{R}_{x_1,l-1}, \dots, \tilde{R}_{x_d,l-1})$. When $\tilde{R}_{x_2,l-1} = \dots = \tilde{R}_{x_d,l-1} = 1$, the function can be simplified as

$$\tilde{R}_{x,l} = f(\tilde{R}_{x_1,l-1}, 1, \dots, 1) = \frac{M \tilde{R}_{x_1,l-1}}{\pi M \tilde{R}_{x_1,l-1}} = M \tilde{R}_{x_1,l-1}.$$

Iterating the function m times, we can write $\tilde{R}_{x,l} = f^{(m)}(\tilde{R}_{y_1,l-m}, \dots, \tilde{R}_{y_{d^m},l-m})$ for some function $f^{(m)} : \Delta_{[k]}^{d^m} \rightarrow \Delta_{[k]}$. A similar calculation shows when $\tilde{R}_{y_2,l-m} = \dots = \tilde{R}_{y_{d^m},l-m} = 1$,

$$\tilde{R}_{x,l} = f^{(m)}(\tilde{R}_{y_1,l-m}, 1, \dots, 1) = M^m \tilde{R}_{y_1,l-m}.$$

Since $f^{(m)}$ is smooth, for some $C_1 = C_1(d, m, M)$ we have that

$$\|\tilde{R}_{x,l} - 1 - \sum_{i=1}^{d^m} (M^m \tilde{R}_{y_i,l-m} - 1)\| \leq C_1 \sum_{i=1}^{d^m} \|\tilde{R}_{y_i,l-m} - 1\|^2 \leq C_1 k \sum_{i=1}^{d^m} \|\tilde{R}_{y_i,l-m} - 1\|^2.$$

By the ergodicity of M , for sufficiently large m and all $\tilde{R} \in \Delta_{[k]}$ we have $\|M^m \tilde{R} - 1\|_\infty \leq \frac{1}{4} \|\tilde{R} - 1\|_\infty$. Therefore for some $\epsilon_1 = \epsilon_1(C_1, k)$ if $R_{y_i,l-m} \leq \epsilon_1$ for all vertices $y_i \in L_{x,m}$ then

$$\|\tilde{R}_{x,l} - 1\|_\infty \leq \left(\frac{1}{4} + C_1 k \epsilon_1\right) \sum_{i=1}^{d^m} \|\tilde{R}_{y_i,l-m} - 1\|_\infty \leq \frac{1}{2} \sum_{i=1}^{d^m} R_{y_i,l-m}. \quad (3.4)$$

This suffices provided that there are no large $R_{y_i,l-m}$.

We now consider the case when there is one large $R_{y_i,l-m}$, which we can without loss of generality assume is $i = 1$. Again since $f^{(m)}$ is smooth, there exists $C_2, \epsilon_2 > 0$ such that for all $\tilde{R}_{y_1,l-m} > \epsilon_1$, if $\sup_{i \geq 2} R_{y_i,l-m} \leq \epsilon_2$ then

$$\|\tilde{R}_{x,l} - M^m \tilde{R}_{y_1,l-m}\| \leq C_2 \sum_{i=2}^{d^m} \|\tilde{R}_{y_i,l-m} - 1\|.$$

Let $\epsilon = \epsilon_2 \wedge (4C_2 d^m k)^{-1} \epsilon_1$, if we moreover have $\sup_{i \geq 2} R_{y_i, l-m} \leq \epsilon$, then

$$\|\tilde{R}_{x,l} - 1\|_\infty \leq \frac{1}{4} \|\tilde{R}_{y_1, l-m} - 1\|_\infty + C_2 d^m k \epsilon \leq \frac{1}{4} R_{y_1, l-m} + \frac{1}{4} \epsilon_1 \leq \frac{1}{2} R_{y_1, l-m}. \quad (3.5)$$

Combining equations (3.4) and (3.5) and noting $\epsilon < \epsilon_1$ completes the proof. \square

So far we have not used the assumption of non-reconstruction and $d\lambda^2 < 1$. In [6], Janson and Mossel develop a notion called robust reconstruction and show that if M satisfies $d\lambda^2 < 1$ then there exist constant $C_1 > 0$ depending on $d\lambda^2$ and $\delta > 0$ depending on d and C_1 such that if for some l , $d_{TV}(\mu_{L_l}^c, \mu_{L_l}) \leq \delta$ for all $c \in [k]$, then $d_{TV}(\mu_{L_{l+1}}^c, \mu_{L_{l+1}}) \leq e^{-C_1} d_{TV}(\mu_{L_l}^c, \mu_{L_l})$. In our case, the existence of such l is guaranteed by the definition of non-reconstruction, hence we have for some constant C_2 that

$$d_{TV}(\mu_{L_l}^c, \mu_{L_l}) \leq C_2 e^{-C_1 l}.$$

A duality argument shows

$$\begin{aligned} \mathbb{E}_{\tau \sim \mu} |\tilde{R}_{x,l}(c) - 1| &= \mathbb{E}_{\tau \sim \mu} \left| \frac{1}{\pi_c} \mu(\sigma_x = c \mid \sigma_{L_{x,l}} = \tau_{L_{x,l}}) - 1 \right| = \mathbb{E}_{\tau \sim \mu} \left| \frac{\mu^c(\sigma_{L_{x,l}} = \tau_{L_{x,l}})}{\mu(\sigma_{L_{x,l}} = \tau_{L_{x,l}})} - 1 \right| \\ &= \sum_{\tau} |\mu^c(\sigma_{L_{x,l}} = \tau_{L_{x,l}}) - \mu(\sigma_{L_{x,l}} = \tau_{L_{x,l}})| = 2d_{TV}(\mu_{L_l}^c, \mu_{L_l}) \leq 2C_2 e^{-C_1 l}. \end{aligned}$$

Maximizing over c we get $\mathbb{E}_{\tau \sim \mu} R_{x,l} \leq C_2 e^{-C_1 l}$ for some (different) constant $C_1, C_2 > 0$ and by Markov's inequality for all $l \geq 1, z > 0$,

$$\Pr_{\tau \sim \mu}(R_{x,l} > z) \leq \frac{C_2}{z} e^{-C_1 l}. \quad (3.6)$$

Proof of Theorem 2.3. By Lemma 3.2, the event $R_{x,l} > z$ implies that either there exist two i such that $R_{y_i, l-m} > \epsilon$ or $\sum_{i=1}^{d^m} R_{y_i, l-m} > 2z$. In the second case if the event $\sum_{i=1}^{d^m} R_{y_i, l-m} > 2z$ holds and for every y_i , $R_{y_i, l-m} \leq \frac{3}{2}z$, then there must exist at least two i such that $R_{y_i, l-m} > \frac{1}{2d^m}z$, otherwise $\sum_{i=1}^{d^m} R_{y_i, l-m} \leq \frac{3}{2}z + \frac{d^m-1}{2d^m}z < 2z$. Therefore we can write

$$\begin{aligned} \Pr_{\tau \sim \mu}(R_{x,l} > z) &\leq \Pr_{\tau \sim \mu}(\exists \text{ two } y_i \in L_{x,m}, R_{y_i, l-m} > \epsilon) + \Pr_{\tau \sim \mu}(\exists y_i \in L_{x,m}, R_{y_i, l-m} > \frac{3}{2}z) \\ &\quad + \Pr_{\tau \sim \mu}(\exists \text{ two } y_i \in L_{x,m}, R_{y_i, l-m} > \frac{1}{2d^m}z). \end{aligned}$$

Let $g(z, l) = \Pr_{\tau \sim \mu}(R_{x,l} > z)$ and $C = \max\{2d^m, \frac{1}{\epsilon \pi_{\min}}\}$, note $g(z, l)$ is a decreasing function in z , the equation

above become

$$\begin{aligned} g(z, l) &\leq d^{2m} g^2(\epsilon, l - m) + d^m g\left(\frac{3}{2}z, l - m\right) + d^{2m} g^2\left(\frac{1}{2d^m}z, l - m\right) \\ &\leq d^m g\left(\frac{3}{2}z, l - m\right) + 2d^{2m} g^2\left(\frac{1}{C}z, l - m\right). \end{aligned}$$

Iterating this estimation h times, we have

$$g(z, l) \leq \sum_{i=0}^h (2d^{2m})^{2^{h-i}(1+i)} g^{2^{h-i}} \left(\left(\frac{3}{2}\right)^i \left(\frac{1}{C}\right)^{h-i} z, l - hm \right). \quad (3.7)$$

where the coefficient can be shown by induction on h using inequality $(a + b)^2 \leq 2(a^2 + b^2)$.

Since for all $z > \pi_{\min}^{-1}$ we have $g(z, l) = 0$, the summand on the RHS of (3.7) is zero for large i . Fix $\kappa = \log \frac{4}{3} C / \log \frac{3}{2} C < 1$, for $h \geq \log(\frac{1}{z\pi_{\min}}) / \log(\frac{4}{3})$ and $i > \kappa h$, we have $(\frac{3}{2})^i (\frac{1}{C})^{h-i} z > \pi_{\min}^{-1}$. Therefore

$$g(z, l) \leq \sum_{i=0}^{\kappa h} (2d^{2m})^{2^{h-i}(1+i)} g^{2^{h-i}} \left(\left(\frac{3}{2}\right)^i \left(\frac{1}{C}\right)^{h-i} z, l - hm \right) \leq \kappa h [(2d^{2m})^h g(C^{-h}z, l - hm)]^{2^{(1-\kappa)h}}$$

Now applying the bound in (3.6) and let $h = rl/m$ for small $r > 0$ such that $(1 - r)C_1 - r \cdot \frac{1}{m} \log(2Cd^{2m}) > \frac{1}{2}C_1 > 0$, for large enough l such that $\log l \leq 2^{\frac{(1-\kappa)r}{m}l}$ we have

$$\begin{aligned} g(z, l) &\leq \kappa h \left((2d^{2m})^h \frac{C_2 C^h}{z} e^{-C_1(l-hm)} \right)^{2^{(1-\kappa)h}} \\ &\leq \frac{\kappa r}{m} l \left(\frac{C_2}{z} (2Cd^{2m})^{\frac{r}{m}l} e^{-C_1(1-r)l} \right)^{2^{\frac{(1-\kappa)r}{m}l}} \leq \frac{\kappa r}{m} \left(\frac{2C_2}{z} e^{-\frac{1}{2}C_1 l} \right)^{2^{\frac{(1-\kappa)r}{m}l}}. \end{aligned}$$

Let $C_3 = \frac{2}{C_1}$, $C_4 = \frac{\kappa r}{m}$, $C_5 = \frac{(1-\kappa)r}{m} \log 2$, for $l > C_3(1 + \log 2C_2 - \log z)$, we have $g(z, l) \leq C_4 \exp\{-\exp(C_5 l)\}$.

Finally define $\xi = \frac{1}{2} \min\{C_3^{-1}, C_5\}$, plug in $z_l = \exp(-\xi l)$. When l is large enough, we have $C_3(1 + \log 2C_2 - \log z) \leq C_3(1 + \log 2C_2) + \frac{1}{2}l < l$ and $\exp(\exp(\frac{1}{2}C_5 l)) > C_4$, therefore

$$\Pr_{\tau \sim \mu} (R_{x,l}(\tau) \geq e^{-\xi l}) = g(z_l, l) \leq C_4 \exp(-e^{C_5 l}) \leq \exp(-e^{\xi l}),$$

completing the proof. \square

4 Proof of Lemma 2.4

The proof of Lemma 2.4 uses a two step analysis. First for block $B_{x,l}$ with sufficiently large l , we consider the measure $\mu_{B_{x,l}}^{*,\tau}$ induced on the upper half of block $B_{x,l/2}$. Denote the following subset of $\Omega_{B_{x,l}}^\tau$,

$$A_\tau = \{\sigma \in \Omega_{B_{x,l}}^\tau, \forall x \in L_{x,l/2+2}, x \text{ is free w.r.t. } \sigma\}.$$

A_τ can be considered as the set of “good” configurations with boundary condition τ . As we will show later, under connectivity condition \mathcal{C} , $\mu_{B_{x,l}}^{*,\tau}(A_\tau)$ is close to 1 with high probability. And as the following lemma claims, conditioning on A_τ and the configuration on boundary $L_{x,l/2}$, the measure induced by $\mu_{B_{x,l}}^{*,\tau}$ actually equals to μ^c . Therefore, as a second step we can apply the result of Lemma 2.3 to $B_{x,l/2}$. Let $\Omega_{L_{x,l/2}}$ be the set of possible configuration on $L_{x,l/2}$.

Lemma 4.1. *Suppose M satisfies the connectivity condition \mathcal{C} . For arbitrary $\tau \in \Omega_{T_x}^c$, $\eta \in \Omega_{L_{x,l/2}}$ and state $c' \in [k]$ that is compatible with c ,*

$$\mu_{B_{x,l}}^{*,\tau}(\sigma_x = c' \mid \sigma_{L_{x,l/2}} = \eta, \sigma \in A_\tau) = \mu^c(\sigma_x = c' \mid \sigma_{L_{x,l/2}} = \eta). \quad (4.1)$$

Proof. For convenience of notation, abbreviate $\sigma_{(1)} = \sigma_{B_{x,l/2-1}}$, $\sigma_{(2)} = \sigma_{B_{x,l} \setminus B_{x,l/2}}$, so every configuration $\sigma \in \Omega_{B_{x,l}}$ can be written as a three tuple $(\sigma_{(1)}, \eta, \sigma_{(2)})$. We of course have that $\sigma_{(1)}, \sigma_{(2)}$ are conditionally independent given $\sigma_{L_{x,l/2}} = \eta$. By the definition of A_τ , $\{\sigma \in A_\tau\}$ only depends on $\sigma_{(2)}$. Therefore to show (4.1), it is enough to show that conditioned on $\sigma_{L_{x,l/2}}$ and $\sigma \in A_\tau$, $\sigma \sim \tau$ is independent of $\sigma_{(1)}$. From there we have

$$\begin{aligned} \mu_{B_{x,l}}^{*,\tau}(\sigma_x = c' \mid \sigma_{L_{x,l/2}} = \eta, \sigma \in A_\tau) &= \mu^c(\sigma_x = c' \mid \sigma_{L_{x,l/2}} = \eta, \sigma \sim \tau, \sigma \in A_\tau) \\ &= \mu^c(\sigma_x = c' \mid \sigma_{L_{x,l/2}} = \eta, \sigma \in A_\tau) = \mu^c(\sigma_x = c' \mid \sigma_{L_{x,l/2}} = \eta). \end{aligned}$$

Since “ \sim ” is a transitive relation, the conditional independence of $\sigma \sim \tau$ and $\sigma_{(1)}$ follows from the following claim. □

Claim 4.2. *Suppose M satisfies the connectivity condition \mathcal{C} . For each $\tau \in \Omega_{T_x}^c$ and $\eta \in \Omega_{L_{x,l/2}}$, and for all $\sigma = (\sigma_{(1)}, \eta, \sigma_{(2)})$, $\sigma' = (\sigma'_{(1)}, \eta, \sigma_{(2)}) \in \Omega_{B_{x,l}}^\tau$ if $\sigma, \sigma' \in A_\tau$, then $\sigma \sim \sigma'$.*

Proof. By Lemma 2.1, there exist a path Γ connecting $\sigma_{(1)}$ to $\sigma'_{(1)}$ in $\Omega_{B_{x,l/2}}^c$, i.e. the Glauber dynamics on $B_{x,l/2}$ with free boundary condition on leaves and the state of $p(x)$ fixed to be c . We will construct a path Γ' in $\Omega_{B_{x,l}}^\tau$ connecting σ to σ' by adding steps between steps of Γ which only changes the configuration on $B_{x,l} \setminus B_{x,l/2}$, such that vertices in $L_{x,l/2+1}$ won't block the move in Γ and after finishing Γ , we can change the configuration on $B_{x,l} \setminus B_{x,l/2}$ back to the original $\sigma_{(2)}$. The construction of Γ' is specified below:

(1) **Before starting Γ .** For each $y \in T$, let $p(y)$ denote the parent of y . For each $y \in L_{x,l/2+2}$, $\sigma \in A_\tau$ implies

that there exist a path Γ_y in T_y changing y from σ_y to $\sigma_{p(p(y))} = \eta_{p(p(y))}$ in one step. To see Γ_y is also a connected path in $B_{x,l}$, we have to show that the parent of y won't block Γ_y . The only neighbor of $p(y)$ in T_y is y and the only move involving y in Γ_y is the last step changing y from σ_y to $\sigma_{p(p(y))}$. The value of $p(y)$ won't block this last step because $\sigma_{p(y)}$ is compatible with both σ_y and $\sigma_{p(p(y))}$ (they are states of neighbouring vertices in σ). Now we will follow each Γ_y for all $y \in L_{x+l/2+2}$ and change σ_y to $\sigma_{p(p(y))}$. After that, for each $w \in L_{x,l/2}$, all vertices in $L_{w,2}$ are in state $\sigma_w = \eta_w$. The configuration on and below $L_{x,l/2+2}$ will henceforth remain fixed until we finish Γ .

(2) **Performing Γ .** For each step in Γ , the existence of $B_{x,l-1} \setminus B_{x,l/2}$ might block this move only if it changes the state of some vertex $w \in L_{x,l/2}$. Suppose it changes w from c_1 to c_2 , remember in the construction above, all vertices in $L_{w,2}$ have states η_w . By part 1 of \mathcal{C} , we can find $c_3 \in [k]$ which is compatible with c_1, c_2 and η_w . Now in order to change w from c_1 to c_2 , it suffices to first change the state of every vertex $z \in L_{w,1}$ to c_3 , and then change w from c_1 to c_2 . This construction keeps the configuration on and below $L_{x,l/2+2}$ unchanged.

(3) **After Γ .** The configuration in $B_{x,l/2}$ is $(\sigma'_{(1)}, \eta)$. We can change every vertex $z \in L_{x,l/2+1}$ back to $\sigma'_z = \sigma_z$ because at this moment its parent $p(z) \in L_{x,l/2}$ and all children of z in $L_{z,1}$ have state $\eta_{p(z)} = \sigma_{p(z)}$, which is compatible with σ_z . From there, we can reverse the path Γ_y for each $y \in L_{x,l/2+2}$ and change the configuration on and below $L_{x,l/2+2}$ back to the original configuration $\sigma_{(2)}$. This completes the construction achieving $\sigma'_{(2)} = \sigma_{(2)}$. \square

Note Claim 4.2 combined with Lemma 4.3 below implies that, with high probability (i.e. on A_τ), the fixed boundary Glauber dynamics on $B_{x,l/2}$ is actually connected as a subgraph of the Glauber dynamic on larger block $B_{x,l}$. This is one part of the proof where connectivity condition is used. We may replace the present connectivity condition by a more general assumption that the probability that the fixed boundary Glauber dynamics on $B_{x,l}$ is not connected in a larger block $B_{x,l'}$ decays double exponentially fast in block size l .

Lemma 4.3. *The connectivity condition \mathcal{C} implies that there exist constants $C_1 > 1$, $C_2 > 0$ such that for all $l \geq 1$,*

$$1 - p_l^{\text{free}} \leq C_2 \exp(-C_1^l). \quad (4.2)$$

Proof. Fix $x \in T$ and $\sigma \in \Omega_{T_x}$. First if for all $1 \leq i \leq d$, $z_i \in L_{x,1}$ is free, then x is also free. To see that, for any $c \in [k]$, by connectivity condition there exists $c' \in [k]$ such that c' is compatible with both c and σ_x , we can first change all z_i to c' in one step and then change x from σ_x to c as the final step.

Now consider $y_{ij} \in L_{z_i,1} \subset L_{x,2}$ for $1 \leq i, j \leq d$. If at most one of the y_{ij} 's is not free, say $y_{11} \in L_{z_1,1}$, then for $i \neq 1$, z_i is free and z_1 can change in one step to all states compatible with $\sigma_{y_{11}}$. Again by \mathcal{C} , for all $c \in [k]$ there exists $c' \in [k]$ such that c' is compatible with c , σ_x and $\sigma_{y_{11}}$. By the construction above, we can change x from σ_x to c in one step, hence x is also free.

This implies if x is not free, then there exist at least two $y_{ij} \in L_{x,2}$ that are not free. By part 2 of \mathcal{C} , there

exists $l_0 > 0$, such that for all $l > l_0$ we have $1 - p_l^{\text{free}} < \frac{1}{d^8}$ and hence

$$1 - p_l^{\text{free}} \leq \binom{d^2}{2} (1 - p_{l-2}^{\text{free}})^2 \leq d^4 (1 - p_{l-2}^{\text{free}})^2 \leq (1 - p_{l-2}^{\text{free}})^{1.5}.$$

By induction, $1 - p_l^{\text{free}} \leq (1 - p_{l_0}^{\text{free}})^{(1.5)^{(l-l_0)/2}}$ which completes the proof. \square

Now we can finish the proof of Lemma 2.4 from which Theorem 1.2 follows immediately.

Proof of Lemma 2.4. Let $C = \alpha(l/2 + 1 - \delta)^2 / [(1 - \delta)^2 \mu^c(\sigma_x = c')]$ be the number on the left hand side of (2.2). It is enough to show for some constant $l_1 \geq 2l_0$, and some coefficient $K \geq 1$, for all $l \geq l_1$

$$\Pr_{\tau \sim \mu^c} \left(|\mu^c(\sigma_x = c' \mid \sigma \sim \tau) - \mu^c(\sigma_x = c')| \geq \frac{K}{C} \right) \leq e^{-2C/K}.$$

To see the sufficiency, note this is just equation (2.1) with δ_1 satisfying $1 - \delta_1 = \frac{1}{4K}(1 - \delta)$.

Recall $A_\tau = \{\sigma \in \Omega_{B_{x,l}}^\tau : \forall x \in L_{x,l/2+2}, x \text{ is free in } \sigma\}$. Lemma 4.3 implies that for some constant $C_1 > 1$, $C_2 > 0$, and $l \geq 1$

$$\begin{aligned} \mathbb{E}_{\tau \sim \mu^c}(\mu_{B_{x,l}}^{*,\tau}(A_\tau^c)) &= \mathbb{E}_{\tau \sim \mu^c}(\mu^c(\sigma \notin A_\tau \mid \sigma \sim \tau)) \\ &= \Pr_{\sigma \sim \mu^c}(\exists y \in L_{x,l/2+2}, y \text{ is not free}) \\ &\leq d^{l/2+2} (1 - p_{l/2-2}^{\text{free}}) \leq C_2 d^{l/2+2} \exp(-C_1^{l/2-2}). \end{aligned}$$

By Markov inequality,

$$\Pr_{\tau \sim \mu^c} \left(\mu_{B_{x,l}}^{*,\tau}(A_\tau^c) > \frac{1}{2C} \right) \leq 2C \mathbb{E}_{\tau \sim \mu^c}(\mu_{B_{x,l}}^{*,\tau}(A_\tau^c)) \leq C d^{l/2+2} C_2 \exp(-C_1^{l/2-2}) \rightarrow 0, \quad (4.3)$$

as $l \rightarrow \infty$. On the event $\{\tau : \mu_{B_{x,l}}^{*,\tau}(A_\tau^c) \leq \frac{1}{2C}\}$,

$$\begin{aligned} \mu_{B_{x,l}}^{*,\tau}(\sigma_x = c' \mid \sigma \in A_\tau) &= \frac{\mu_{B_{x,l}}^{*,\tau}(\sigma_x = c', \sigma \in A_\tau)}{\mu_{B_{x,l}}^{*,\tau}(\sigma \in A_\tau)} \leq \frac{1}{1 - \frac{1}{2C}} \mu_{B_{x,l}}^{*,\tau}(\sigma_x = c', \sigma \in A_\tau) \\ &\leq (1 + \frac{1}{C}) \left(\mu_{B_{x,l}}^{*,\tau}(\sigma_x = c') + \frac{1}{C} \right) \leq \mu_{B_{x,l}}^{*,\tau}(\sigma_x = c') + \frac{3}{C} \\ \mu_{B_{x,l}}^{*,\tau}(\sigma_x = c' \mid \sigma \in A_\tau) &\geq \mu_{B_{x,l}}^{*,\tau}(\sigma_x = c', \sigma \in A_\tau) \geq \mu_{B_{x,l}}^{*,\tau}(\sigma_x = c') - \frac{1}{C}. \end{aligned} \quad (4.4)$$

Combining the two results together we have

$$\left| \mu_{B_{x,l}}^{*,\tau}(\sigma_x = c') - \mu_{B_{x,l}}^{*,\tau}(\sigma_x = c' \mid \sigma \in A_\tau) \right| \leq \frac{3}{C}. \quad (4.5)$$

Now splitting $\mu_{B_{x,l}}^{*,\tau}(\sigma_x = c' \mid \sigma \in A_\tau)$ according to $\sigma_{L_{x,l/2}}$ and applying Lemma 4.1, we have

$$\begin{aligned}\mu_{B_{x,l}}^{*,\tau}(\sigma_x = c' \mid \sigma \in A_\tau) &= \sum_{\eta} \mu_{B_{x,l}}^{*,\tau}(\sigma_x = c' \mid \sigma \in A_\tau, \sigma_{L_{x,l/2}} = \eta) \mu_{B_{x,l}}^{*,\tau}(\sigma_{L_{x,l/2}} = \eta \mid \sigma \in A_\tau) \\ &= \sum_{\eta} \mu^c(\sigma_x = c' \mid \sigma_{L_{x,l/2}} = \eta) \mu_{B_{x,l}}^{*,\tau}(\sigma_{L_{x,l/2}} = \eta \mid \sigma \in A_\tau).\end{aligned}\quad (4.6)$$

We would like to estimate the set of η such that $\mu^c(\sigma_x = c' \mid \sigma_{L_{x,l/2}} = \eta)$ has a large bias. Let $B = \{\eta : |\mu^c(\sigma_x = c' \mid \sigma_{L_{x,l/2}} = \eta) - \mu^c(\sigma_x = c')| \geq \frac{1}{C}\}$, Theorem 2.3 implies that for $l/2 \geq l_0$ and some $\delta > 0$, we have $\Pr_{\eta \sim \mu^c}(B) \leq e^{-2C}$, where $\eta \sim \mu^c$ means the induced measure on $L_{x,l/2}$. Again by Markov's inequality,

$$\Pr_{\tau \sim \mu^c}(\mu_{B_{x,l}}^{*,\tau}(\sigma_{L_{x,l/2}} \in B) > \frac{1}{C}) \leq C \mathbb{E}_{\tau \sim \mu^c} \mu_{B_{x,l}}^{*,\tau}(\sigma_{L_{x,l/2}} \in B) = C \mu^c(B) \leq C e^{-2C}. \quad (4.7)$$

On the event $\{\tau : \mu_{B_{x,l}}^{*,\tau}(\sigma_{L_{l/2}} \in B) \leq \frac{1}{C}\} \cap \{\tau : \mu_{B_{x,l}}^{*,\tau}(A_\tau^c) \leq \frac{1}{2C}\}$, from (4.6) we have

$$\begin{aligned}\left| \mu_{B_{x,l}}^{*,\tau}(\sigma_x = c' \mid \sigma \in A_\tau) - \mu^c(\sigma_x = c') \right| &\leq \sum_{\eta} \left| \mu^c(\sigma_x = c' \mid \sigma_{L_{l/2}} = \eta) - \mu^c(\sigma_x = c') \right| \mu_{B_{x,l}}^{*,\tau}(\sigma_{L_{l/2}} = \eta \mid \sigma \in A_\tau) \\ &\leq \sum_{\eta \in B^c} \frac{1}{C} \mu_{B_{x,l}}^{*,\tau}(\sigma_{L_{l/2}} = \eta \mid \sigma \in A_\tau) + \mu_{B_{x,l}}^{*,\tau}(\sigma_{L_{l/2}} \in B \mid \sigma \in A_\tau) \\ &\leq \frac{1}{C} \cdot 1 + \frac{1}{C} + \frac{3}{C} = \frac{5}{C}\end{aligned}\quad (4.8)$$

where the last inequality follows from similar argument to (4.4).

Combining the result of equations (4.5) and (4.8), on the event $\{\tau : \mu_{B_{x,l}}^{*,\tau}(\sigma_{L_{l/2}} \in B) \leq \frac{1}{C}\} \cap \{\tau : \mu_{B_{x,l}}^{*,\tau}(A_\tau^c) \leq \frac{1}{2C}\}$, we have

$$\left| \mu_{B_{x,l}}^{*,\tau}(\sigma_x = c') - \mu^c(\sigma_x = c') \right| \leq \frac{3}{C} + \frac{5}{C} = \frac{8}{C}.$$

Therefore using the bounds from (4.3) and (4.7), for all $l \geq 2l_0$,

$$\begin{aligned}\Pr_{\tau \sim \mu^c} \left(\left| \mu^c(\sigma_x = c' \mid \sigma \sim \tau) - \mu^c(\sigma_x = c') \right| > \frac{8}{C} \right) &\leq \Pr(\mu_{B_{x,l}}^{*,\tau}(\sigma_{L_{l/2}} \in B) \leq \frac{1}{C}) + \Pr(\mu_{B_{x,l}}^{*,\tau}(A_\tau^c) \leq \frac{1}{2C}) \\ &\leq C d^{l/2+2} C_2 \exp(-C_1^{l/2-2}) + C e^{-2C} \leq e^{-16C}.\end{aligned}$$

where recall that $C \approx C' l^2$ for some constant C' depending on δ and α , the last step is true for some large enough constant \tilde{l} depending on d, C_1, C_2 and C' . This means the strong concentration inequality (2.1) holds, for $K = 8$, $\delta_1 = 1 - \frac{1}{4K}(1 - \delta)$ and $l_1 = \max\{2l_0, \tilde{l}\}$. Moreover, by taking l large enough and changing the constant C to $8C$ in (4.5) and (4.8), we can make K arbitrarily close to 1. \square

5 Component dynamics version of fast mixing results

In this section we prove Theorem 2.2. The theorem was originally proved for block dynamics in [10]. Here we give a modification of their theorem adapted to the component dynamics by roughly “adding stars” at all occurrence of $B_{x,l}$. We will only state the key steps and refer the details to [10]. For the remainder of this section, we let $\mu = \mu_T^c, \Omega = \Omega_T^c$. Recall that $\tilde{T}_x = T_x \setminus \{x\}$. First we define the entropy mixing condition for Gibbs measure to be the following:

Definition (Entropy Mixing). We say that μ satisfies $\text{EM}^*(l, \epsilon)$ if for every $x \in T, \eta \in \Omega$ and any $f \geq 0$ that does not depend on the connected component of $B_{x,l}$, i.e. $f(\sigma) = \mu_{B_{x,l}}^{*,\sigma}(f), \forall \sigma \in \Omega$, we have $\text{Ent}_{T_x}^\eta[\mu_{\tilde{T}_x}(f)] \leq \epsilon \cdot \text{Ent}_{T_x}^\eta(f)$ where $\text{Ent}_{T_x}^\eta$ means the entropy w.r.t $\mu_{T_x}^\eta$.

Denote $p_{\min} = \min_{c,c' \in [k]} \{M(c, c') : M(c, c') > 0\}$. By the Markov chain construction of configurations, we have $p_{\min} = \min_{x,c,c'} \{\mu_{T_x}^c(\sigma_x = c'), c, c' \text{ are compatible}\}$. The following theorem relates entropy mixing condition to the log-Sobolev constant.

Theorem 5.1. *For any l and $\delta > 0$, if μ satisfies $\text{EM}^*(l, [(1-\delta)p_{\min}/(l+1-\delta)]^2)$ then $\text{Ent}(f) \leq \frac{2}{\delta} \cdot \mathcal{E}_l^*(f)$.*

To show Theorem 5.1, we need the following modification of Lemma 3.5 (ii) of [10]. The proof follows from its analog in [10] immediately once we replace $\nu_A, \text{Ent}_A, \nu_B, \text{Ent}_B$ there with $\nu_{\tilde{T}_x}, \text{Ent}_{\tilde{T}_x}, \nu_{B_{x,l}}^*, \text{Ent}_{B_{x,l}}^*$ respectively.

Lemma 5.2. *For any $\epsilon < p_{\min}^2$, if μ satisfies $\text{EM}^*(l, \epsilon)$ then for every $x \in T$, any $\eta \in \Omega$ and any $f \geq 0$ we have $\text{Ent}_{T_x}^\eta[\mu_{\tilde{T}_x}(f)] \leq \frac{1}{1-\epsilon'} \cdot \mu_{T_x}^\eta[\text{Ent}_{B_{x,l}}^*(f)] + \frac{\epsilon'}{1-\epsilon'} \cdot \mu_{T_x}^\eta[\text{Ent}_{\tilde{T}_x}(f)]$ with $\epsilon' = \sqrt{\epsilon}/p_{\min}$.*

Now plugging $\epsilon = [(1-\delta)p_{\min}/(l+1-\delta)]^2$ into Lemma 5.2 verifies the hypothesis of the following claim, which then implies Theorem 5.1:

Claim 5.3. *If for every $x \in T, \eta \in \Omega$ and any $f \geq 0$,*

$$\text{Ent}_{T_x}^\eta[\mu_{\tilde{T}_x}(f)] \leq c \cdot \mu_{T_x}^\eta[\text{Ent}_{B_{x,l}}^*(f)] + \frac{1-\delta}{l} \cdot \mu_{T_x}^\eta[\text{Ent}_{\tilde{T}_x}(f)], \quad (5.1)$$

then $\text{Ent}(f) \leq \frac{c}{\delta} \cdot \mathcal{E}_l^(f)$ for all $f \geq 0$.*

Proof. First we decompose $\text{Ent}(f)$ as a sum of $\text{Ent}_{T_x}^\eta[\mu_{\tilde{T}_x}(f)]$. Suppose T have m levels, consider $\emptyset = F_0 \subset F_1 \subset \dots \subset F_{m+1} = T$, where F_i is the lowest i levels of T . By basic properties of conditional entropy (equation (3), (4), (5) of [10]) and Markov property of Gibbs measure, we have

$$\text{Ent}(f) = \dots = \sum_{i=1}^{m+1} \mu[\text{Ent}_{F_i}(\mu_{F_{i-1}}(f))] \leq \sum_{i=1}^{m+1} \sum_{x \in F_i \setminus F_{i-1}} \mu[\text{Ent}_{T_x}(\mu_{F_{i-1}}(f))] \leq \sum_{x \in T} \mu[\text{Ent}_{T_x}(\mu_{\tilde{T}_x}(f))]. \quad (5.2)$$

Denote the final sum by $\text{PEnt}(f)$. For each term in the sum of $\text{PEnt}(f)$, apply (5.1) to $g = \mu_{T_x \setminus B_{x,l} \cup \partial B_{x,l}}(f)$ and perform the decomposition trick of (5.2) again, we have for every $x \in T$ and $\eta \in \Omega$.

$$\begin{aligned} \text{Ent}_{T_x}^\eta[\mu_{\tilde{T}_x}(f)] &= \text{Ent}_{T_x}^\eta[\mu_{\tilde{T}_x}(g)] \leq c \cdot \mu_{T_x}^\eta[\text{Ent}_{B_{x,l}}^*(g)] + \frac{1-\delta}{l} \cdot \mu_{T_x}^\eta[\text{Ent}_{\tilde{T}_x}(g)] \\ &\leq c \cdot \mu_{T_x}^\eta[\text{Ent}_{B_{x,l}}^*(f)] + \frac{1-\delta}{l} \cdot \sum_{y \in B_{x,l} \cup \partial B_{x,l}, y \neq x} \mu_{T_x}^\eta[\text{Ent}_{T_y}(\mu_{\tilde{T}_y}(f))]. \end{aligned}$$

Now sum up for all $x \in T$ and take expectation w.r.t μ for $\eta \in \Omega$, noting that the first term of the last line sums up to $\mathcal{E}_l^* = \sum_{x \in T} \mu(\text{Ent}_{B_{x,l}}^*(f))$ and each y in second term appears in at most l blocks so we get

$$\begin{aligned} \text{PEnt}(f) &\leq c \cdot \mathcal{E}_l^*(f) + \frac{1-\delta}{l} \cdot \sum_{x \in T} \sum_{y \in B_{x,l} \cup \partial B_{x,l}, y \neq x} \mu[\text{Ent}_{T_y}(\mu_{\tilde{T}_y}(f))] \\ &\leq c \cdot \mathcal{E}_l^*(f) + \frac{1-\delta}{l} \cdot l \cdot \sum_{y \in T} \mu[\text{Ent}_{T_y}(\mu_{\tilde{T}_y}(f))] = c \cdot \mathcal{E}_l^*(f) + (1-\delta) \cdot \text{PEnt}(f), \end{aligned}$$

and hence $\text{Ent}(f) \leq \text{PEnt}(f) \leq \frac{c}{\delta} \cdot \mathcal{E}_l^*$. \square

Given the result of Theorem 5.1, it is enough to show that for some constant α , the super concentration inequality of (2.1) implies $\text{EM}^*(l, [(1-\delta)p_{\min}/(l+1-\delta)]^2)$. For convenience of notation, define following two functions for each $c' \in [k]$:

$$g_{c'}(\sigma) = \frac{\mu(\sigma | \sigma_\rho = c')}{\mu(\sigma)} = \frac{1}{\mu(\sigma_\rho = c')} \cdot 1_{\{\sigma_\rho = c'\}}, g_{c'}^{*(l)} = \mu_{B_{\rho,l}}^*(g_{c'}).$$

Letting $\delta' = (1-\delta)^2/\alpha(l+1-\delta)^2$, we can rewrite (2.1) as

$$\mu\left(\left|g_{c'}^{*(l)} - 1\right| > \delta'\right) \leq e^{-2/\delta'}. \quad (5.3)$$

Theorem 5.4. *There exists a constant C such that if (5.3) holds for some $\delta' \geq 0$ and all pairs of states $c, c' \in [k]$, we have $\text{Ent}[\mu_{\tilde{T}}(f)] \leq C\delta' \text{Ent}(f)$ for any $f \geq 0$ satisfying $f(\sigma) = \mu_{B_{\rho,l}}^{*,\sigma}(f), \forall \sigma \in \Omega^c$, i.e. $\text{EM}^*(l, C\delta')$ holds.*

Proof. Since for any $f' \geq 0$, $\text{Ent}(f') \leq \text{Var}(f')/\mu(f)$, we can write

$$\begin{aligned} \text{Ent}[\mu_{\tilde{T}}(f)] &\leq \frac{\text{Var}[\mu_{\tilde{T}}(f)]}{\mu(\mu_{\tilde{T}}(f))} = \frac{1}{\mu(f)} \sum_{c' \in [k]} \mu(\sigma_\rho = c') (\mu(f | \sigma_\rho = c') - \mu(f))^2 \\ &= \frac{1}{\mu(f)} \sum_{c' \in [k]} \mu(\sigma_\rho = c') \text{Cov}(g_{c'}, f)^2 \leq \max_{c' \in [k]} \frac{\text{Cov}(g_{c'}, f)^2}{\mu(f)} = \max_{c' \in [k]} \frac{\text{Cov}(g_{c'}^{*(l)}, f)^2}{\mu(f)}. \end{aligned} \quad (5.4)$$

where covariance is taken w.r.t. μ and the last step is because $f(\sigma) = \mu_{B_{\rho,l}}^{*,\sigma}(f)$. Now using Lemma 5.4 of [10] (cited below) with

$$f_1 = \frac{g_{c'}^{*(l)} - 1}{\|g_{c'}^{*(l)}\|_\infty}, f_2 = \frac{f}{\mu(f)}$$

and noting that $\|g_{c'}^{*(l)}\|_\infty \leq \|g_{c'}\|_\infty \leq p_{\min}$, we have $\text{Cov}(g_{c'}^{*(l)}, f)^2 \leq C\delta'\mu(f)\text{Ent}(f)$ for some constant $C = C'/p_{\min}^2$. Plug it into (5.4), we get $\text{Ent}[\mu_{\tilde{T}}(f)] \leq C\delta'\text{Ent}(f)$. \square

Lemma 5.5 (Lemma 5.4 of [10]). *Let $\{\Omega, \mathcal{F}, \nu\}$ be a probability space and let f_1 be a mean-zero random variable such that $\|f\|_\infty \leq 1$ and $\nu[|f_1| > \delta] \leq e^{-2/\delta}$ for some $\delta \in (0, 1)$. Let f_2 be a probability density w.r.t ν , i.e. $f_2 \geq 0$ and $\nu(f_2) = 1$. Then there exists a numerical constant $C' > 0$ independent of ν, f_1, f_2 and δ , such that $\nu(f_1 f_2) \leq C'\delta\text{Ent}_\nu(f_2)$.*

Now, as the last step, letting $\alpha = C/p_{\min}^2$ for the constant C in Theorem 5.4 finishes the proof of Theorem 2.2.

6 Results for k -coloring

In this section we prove Theorem 1.1, for which it is enough to verify the connectivity condition \mathcal{C} , in particular to show $p_l^{\text{free}} \rightarrow 1$, as $l \rightarrow \infty$. In fact for the coloring model, as we will show in a moment, a vertex can change to all k states in one step if all its children can change to 2 or 3 states in one step. We will first formalize this idea by defining the “type” of a vertex and then analyze the recursion with this weaker notion.

Recall the definition that for given configuration $\sigma \in \Omega_T$ with $\sigma_\rho = c$, we say the root can change to color c' in one step iff there exist a path $\sigma = \sigma^0, \sigma^1, \dots, \sigma^n \in \Omega_T^\sigma$ such that for each $i, \sigma^i, \sigma^{i+1}$ differs by only one vertex and $\sigma_\rho^i = c, i = 0, 1, \dots, n-1, \sigma_\rho^n = c'$. Let $C(\rho)$ denote the set of colors the root can change to in one step (including the original color). We define the type of root to be rigid/type 2/type 3 if $|C(\rho)| = 1/2/\geq 3$ respectively. For general vertex $x \in T$, not necessarily the root, we can similarly define $C(x)$ and rigid/type 2/type 3 by taking x as the root of subtree T_x and considering $\sigma|_{T_x}$. The set $C(x)$ is a function of σ_{T_x} and is independent of the rest of the tree.

Let $p_l^r = \mu_l(\text{the root is rigid})$, where μ_l is the Gibbs measure on l -level tree with free boundary condition. Define $p_l^{(2)}, p_l^{(3)}$ similarly, we have $p_l^r + p_l^{(2)} + p_l^{(3)} = 1$. For tree T with $l' > l$ levels and $x \in T$ that is l levels above the bottom boundary, noting $\mu_{l'}|_{T_x} = \mu_l$, we also have

$$\mu_{l'}(x \text{ is rigid/type 2/type 3}) = \mu_{l'}|_{T_x}(x \text{ is rigid/type 2/type 3}) = p_l^r/p_l^{(2)}/p_l^{(3)}.$$

The definition above is independent of the parent of x . In order to recursively analysing these probabilities, we need one further definition describing how the type of a vertex affects the type of its parent. For a given configuration $\sigma \in \Omega_T$ and $x \in \tilde{T} = T \setminus \{\rho\}$, recall $p(x)$ is the parent of x , we say x is bad iff $C(x) \setminus \{\sigma_{p(x)}\} = \{\sigma_x\}$.

Otherwise we say x is good which implies $|C(x) \setminus \{\sigma_{p(x)}\}| \geq 2$, i.e. x has at least one more choice other than $\sigma_{p(x)}$. Note that the event that x is bad depends only on $\sigma|_{T_{p(x)}}$ and given σ_x , for $x_i \in L_{x,1}$, events $\{x_i \text{ is bad}\}$ are conditionally i.i.d. and independent of the configurations outside T_x . Hence, by similar argument, we can define $p_l^b = 1 - p_l^g = \mu_{l'}(x \text{ is bad})$. The relation of rigid/type 2/type 3 and good/bad is shown in the following lemma.

Lemma 6.1. *For $l' > l > 0$ and $x \in T$ l levels above the bottom boundary,*

$$\mu_{l'}(x \text{ is bad} \mid x \text{ is rigid})=1, \mu_{l'}(x \text{ is bad} \mid x \text{ is type 2})=\frac{1}{k-1}, \mu_{l'}(x \text{ is bad} \mid x \text{ is type 3})=0, \quad (6.1)$$

Hence $p_l^b = p_l^r + \frac{1}{k-1}p_l^{(2)}, p_l^g = p_l^{(3)} + \frac{k-2}{k-1}p_l^{(2)}$.

Proof. The first and third equation of (6.1) is obvious as $|C(x)|$ and $|C(x) \setminus \{\sigma_{p(x)}\}|$ differs at most by one, and the equality about p_l^b and p_l^g follows immediately from the (6.1). Hence we only need to show the second equation. Given $|C(x)| = 2$, x is bad iff $\sigma_{p(x)} \in C(x)$. So the conditional probability can be written as $P(C(x) = \{\sigma_{p(x)}, \sigma_x\} \mid |C(x)| = 2)$.

Note $C(x)$ only depends on T_x , in particular it is conditionally independent of $\sigma_{p(x)}$ given σ_x . By symmetry, the distribution of $C(x) \setminus \{\sigma_x\}$ given $|C(x)|$ and σ_x is uniformly distribution with probability $1/\binom{k-1}{|C(x)|-1}$. Hence

$$\Pr(C(x) = \{\sigma_{p(x)}, \sigma_x\} \mid |C(x)| = 2) = \frac{1}{\binom{k-1}{1}} = \frac{1}{k-1}.$$

□

The next lemma follows a similar argument to Claim 4.2 and Lemma 4.3, and shows that in order to bound the probability of a vertex being free, it is enough to bound the probability of being bad.

Lemma 6.2. *Suppose $k \geq 4$. For any $\sigma \in \Omega_T$ and $x \in T$, if every child of x is good, then x is free.*

Proof. Fix $c \in [k]$. Then since each child y_i of x is good there exist $c_i \in C(y_i) \setminus \{c, \sigma_x\}$. So to change x from σ_x to c in one step, we can first change the color of every y_i to c_i in one step and then in the final step change x from σ_x to c . Hence x is free. □

Now we will show that for large enough k , in the region of non-reconstruction, the probability of seeing a bad vertex l levels above bottom decays double exponentially fast in l . In fact we will prove the result for a region slightly larger than the known non-reconstruction region, which is $d \leq k[\log k + \log \log k + \beta]$, for any $\beta < 1 - \ln 2$ (see [15]).

Theorem 6.3. Suppose $\beta < 1$, For sufficiently large k and $d \leq k[\log k + \log \log k + \beta]$, there exists a constant l_0 depending only on k and d , such that for $l \geq l_0$,

$$p_l^b \leq \exp(-(k/2)^{l-l_0}). \quad (6.2)$$

Combining Lemma 6.2 and Theorem 6.3, we can now finish the proof of Theorem 1.1.

Proof of Theorem 1.1. By Theorem 1.2 and the known bounds on the reconstruction threshold for colorings [15] it is enough to show that the connectivity condition holds. The first part of the condition is obviously true for $k \geq 4$. For the second condition,

$$1 - p_l^{\text{free}} = \Pr_{\sigma \sim \mu_l}(\text{root is not free}) \leq \Pr(\exists x \in L_1, x \text{ is bad}) \leq dp_{l-1}^b \leq d \exp(-(\frac{1}{2}k)^{l-l_0}),$$

which tends to 0 as l tends to infinity completing the proof. \square

The proof of Theorem 6.3 is split into two phases, when p_l^b is close to 1 and when p_l^b is smaller than $\frac{1}{ed}$.

Lemma 6.4. Under the assumption of Theorem 6.3, there exist a constant l_0 depending only on k and d such that $p_{l_0}^b < \frac{1}{ed}$.

Proof. This proof is similar to Lemma 2 and Lemma 4 of [15]. We recursively analyze the probabilities as a function of the depth of the tree l . For $l = 0$, T consist only the bottom boundary and hence $p_0^r = 1, p_0^{(2)} = p_0^{(3)} = 0, p_0^b = p_0^r + \frac{1}{k-1}p_0^{(2)} = 1$.

For $l \geq 1$, suppose without loss of generality that the color of the root is 1 and its children are $x_1, \dots, x_d \in L_1$. Let \mathcal{F} denote the sigma-field generated by $\{\sigma_{x_i}, i = 1, \dots, d\}$ and let $d_c = \#\{i, \sigma_{x_i} = c\}$ be the number of children with color c for $2 \leq c \leq k$. By definition, the sizes $|C(x_i)|$ and hence the type of x_i are independent of \mathcal{F} and are i.i.d. distributed. Conditioning on \mathcal{F} and $(|C(x_i)|)_{i=1}^d$, the set $C(x_i) \setminus \{\sigma_{x_i}\}$ is uniformly randomly chosen from all subsets of $[k] \setminus \{\sigma_{x_i}\}$ with $(|C(x_i)| - 1)$ elements. Therefore the number of bad vertices of color c given \mathcal{F} is distributed as $\text{Bin}(d_c, p_{l-1}^b)$.

Following similar argument of Lemma 6.2, the root can change to color c in one step if and only if none of the x_i 's with color c is bad, which happens with probability $(1 - p_{l-1}^b)^{d_c}$. Therefore we have

$$p_l^b = p_l^r + \frac{1}{k-1}p_l^{(2)} = \prod_{c=2}^k \mathbb{E}[1 - (1 - p_{l-1}^b)^{d_c}] + \frac{1}{k-1} \sum_{c'=2}^k \mathbb{E}\left[(1 - p_{l-1}^b)^{d_{c'}} \prod_{c \neq c'} (1 - (1 - p_{l-1}^b)^{d_c})\right].$$

Viewing the right hand side as a function of (d_2, \dots, d_k) , increasing d_c means adding more vertices of color c , which increases the probability of blocking the move of the root. Therefore p_l^b is an increasing function w.r.t every d_c . By symmetry, (d_2, \dots, d_k) follows a multi-nominal distribution. Fix $\beta < \beta^* < 1$, let \tilde{d}_c be i.i.d. Poisson(D) random variables where $D = \log k + \log \log k + \beta^*$. We can couple (d_2, \dots, d_k) and $(\tilde{d}_2, \dots, \tilde{d}_k)$ such

that $(d_2, \dots, d_k) \leq (\tilde{d}_2, \dots, \tilde{d}_k)$ whenever $\sum_{c=2}^k \tilde{d}_c \geq d$. Letting $p = \Pr(\text{Poisson}((k-1)D) < d)$, the recursion relationship satisfies

$$\begin{aligned}
p_l^b &= p_l^r + \frac{1}{k-1} p_l^{(2)} = \prod_{c=2}^k \mathbb{E} [1 - (1 - p_{l-1}^b)^{d_c}] + \frac{1}{k-1} \sum_{c'=2}^k \mathbb{E} \left[(1 - p_{l-1}^b)^{d_{c'}} \prod_{c \neq c'} (1 - (1 - p_{l-1}^b)^{d_c}) \right] \\
&\leq \prod_{c=2}^k \mathbb{E} [1 - (1 - p_{l-1}^b)^{d_c}] + \frac{1}{k-1} \sum_{c'=2}^k \mathbb{E} (1 - p_{l-1}^b)^{d_{c'}} \prod_{c \neq c'} \mathbb{E} [1 - (1 - p_{l-1}^b)^{d_c}] + p \\
&= (1 - \exp(-p_{l-1}^b D))^{k-1} + \frac{k-1}{k-1} \exp(-p_{l-1}^b D) (1 - \exp(-p_{l-1}^b D))^{k-2} + p \\
&= (1 - \exp(-p_{l-1}^b D))^{k-2} + p \leq \exp(-(k-2) \exp(-p_{l-1}^b D)) + p
\end{aligned}$$

where the last step follows from the fact $(1-r)^k \leq e^{-kr}$ for $0 < r < 1$.

The rest of the proof resembles the argument of Lemma 3 of [15]. Let $f(x) = \exp(-(k-2) \exp(-xD)) + p$, $f(x)$ is an increasing function in x . So let $y_0 = p_0^b = 1$, $y_l = f(y_{l-1})$, we have $p_l^b \leq y_l$, for all $l \geq 0$. It is enough to show the existence of an l_0 such that $y_{l_0} \leq \frac{1}{ed}$. Since $\frac{d}{dx} \exp(-x) |_{x=0} = -1$, there exist $\epsilon, \delta > 0$ such that for $0 < x < \delta$, $e^{-x} \leq 1 - (1-\epsilon)x$. For large enough k such that $(k-2) \exp(-D) = \frac{k-2}{k \log k} e^{-\beta^*} < \delta$, we have

$$y_1 = f(1) \leq 1 - (1-\epsilon) \frac{k-2}{k \log k} e^{-\beta^*} + p.$$

Since $\beta < \beta^* < 1$, $(k-1)D - d \geq (\beta^* - \beta)k + o(k)$, by Hoeffding's inequality, $p = \exp(-\Omega(\frac{k}{\sqrt{d}})) = o(k^{-2}) = o(d^{-1})$. Therefore, for large enough k ,

$$y_1 \leq 1 - \frac{1-\epsilon}{2e \log k} + o(k^{-1}) \leq 1 = y_0.$$

Hence y_l is a decreasing sequence as long as $(k-2) \exp(-y_l D) < \delta$. Choosing ϵ small enough, there exists $r' > r > 0$ such that $(1-\epsilon)e^{-\beta^*} > e^{-1}(1+r')$, and we have

$$\begin{aligned}
1 - y_{l+1} &\geq 1 - (p + 1 - (1-\epsilon)(k-2) \exp(-y_l D)) \\
&\geq (1-\epsilon) \frac{(k-2)e^{-\beta^*}}{k \log k} \exp((1-y_l) \log k) - p \\
&\geq \frac{k-2}{k} (1-\epsilon) e^{1-\beta^*} (1-y_l) - p \\
&\geq (1+r')(1-y_l) - p \geq (1+r)(1-y_l)
\end{aligned}$$

where the second last inequality follows from inequality $e^x > ex$, and the last inequality follows from that $1 - y_l \geq 1 - y_1 = O(\frac{1}{\log k})$ while $p = o(k^{-2})$. Therefore after a constant number of steps $(k-2) \exp(-y_l D) \geq \delta$. Now let $e^{-\delta} < \alpha' < \alpha < 1$, for some constants α, α' . When k is large enough, $y_{l+1} \leq p + e^{-\delta} < \alpha' < 1$. Then

again for k large enough, $\exp(-y_{l+1}D) \geq \exp(-\alpha'D) \geq \exp(-\alpha \log k) = k^{-\alpha}$. Therefore for k large enough

$$y_{l+2} \leq p + \exp(-(k-2) \exp(-y_{l+1}D)) \leq p + \exp(-\frac{1}{2}k^{1-\alpha}) \leq \frac{1}{ed}.$$

□

After first l_0 levels, we cannot use the same method because the error of Poisson coupling becomes non-negligible; but meanwhile, p_l^b is small enough such that bounding the total number of bad children is enough to finish the proof.

Proof of Theorem 6.3. In order for a vertex to be bad, there must be at least $k-2$ of its children which are bad. Therefore,

$$p_l^b \leq \binom{d}{k-2} (p_{l-1}^b)^{k-2} \leq (dp_{l-1}^b)^{k-2}.$$

Letting l_0 be the constant in Lemma 6.4 we complete the proof by induction on l for $l \geq l_0$. If $l = l_0$, $p_{l_0}^b \leq \frac{1}{ed} \leq \frac{1}{e}$. If p_l^b satisfies (6.2) for $l > l_0$, then for k large enough such that $\log(2k \log k) \leq \frac{1}{6}k$ and $k-2 \geq \frac{3}{4}k$,

$$\begin{aligned} p_{l+1}^b &\leq (dp_l^b)^{k-2} \leq (2k \log k \exp(-(k/2)^{l-l_0}))^{k-2} = \exp(k-2) \left(-(k/2)^{l-l_0} + \log(2k \log k) \right) \\ &\leq \exp\left(-\frac{3}{4}k \cdot \frac{2}{3}\left(\frac{1}{2}\right)^{l-l_0}\right) = \exp(-(k/2)^{l+1-l_0}). \end{aligned}$$

Therefore (6.2) holds for all $l \geq l_0$. □

References

- [1] Noam Berger, Claire Kenyon, Elchanan Mossel, and Yuval Peres. Glauber dynamics on trees and hyperbolic graphs. *Probability Theory and Related Fields*, 131(3):311–340, 2005.
- [2] Nayantara Bhatnagar, Juan Vera, Eric Vigoda, and Dror Weitz. Reconstruction for colorings on trees. *SIAM Journal on Discrete Mathematics*, 25(2):809–826, 2011.
- [3] Jian Ding, Eyal Lubetzky, and Yuval Peres. Mixing time of critical ising model on trees is polynomial in the height. *Communications in Mathematical Physics*, 295(1):161–207, 2010.
- [4] Alan Frieze and Eric Vigoda. A survey on the use of markov chains to randomly sample colourings. *Oxford Lecture Series in Mathematics and its Applications*, 34:53, 2007.
- [5] Leslie Ann Goldberg, Mark Jerrum, and Marek Karpinski. The mixing time of glauber dynamics for coloring regular trees. *Random Structures & Algorithms*, 36(4):464–476, 2010.

- [6] Svante Janson and Elchanan Mossel. Robust reconstruction on trees is determined by the second eigenvalue. *The Annals of Probability*, 32(3B):2630–2649, 2004.
- [7] Johan Jonasson. Uniqueness of uniform random colorings of regular trees. *Statistics & Probability Letters*, 57(3):243–248, 2002.
- [8] Brendan Lucier, Michael Molloy, and Yuval Peres. The glauher dynamics for colourings of bounded degree trees. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 631–645. Springer, 2009.
- [9] Fabio Martinelli. Lectures on glauher dynamics for discrete spin models. In *Lectures on probability theory and statistics*, pages 93–191. Springer, 1999.
- [10] Fabio Martinelli, Alistair Sinclair, and Dror Weitz. Glauber dynamics on trees: boundary conditions and mixing time. *Communications in Mathematical Physics*, 250(2):301–334, 2004.
- [11] Fabio Martinelli, Alistair Sinclair, and Dror Weitz. Fast mixing for independent sets, colorings, and other models on trees. *Random Structures & Algorithms*, 31(2):134–172, 2007.
- [12] Elchanan Mossel. Survey-information flow on trees. *DIMACS series in discrete mathematics and theoretical computer science*, 63:155–170, 2004.
- [13] Elchanan Mossel, Yuval Peres, et al. Information flow on trees. *The Annals of Applied Probability*, 13(3):817–844, 2003.
- [14] Laurent Saloff-Coste. Lectures on finite markov chains. In *Lectures on probability theory and statistics*, pages 301–413. Springer, 1997.
- [15] Allan Sly. Reconstruction of random colourings. *Communications in Mathematical Physics*, 288(3):943–961, 2009.
- [16] Prasad Tetali, Juan C Vera, Eric Vigoda, and Linji Yang. Phase transition for the mixing time of the glauher dynamics for coloring regular trees. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1646–1656. Society for Industrial and Applied Mathematics, 2010.
- [17] Eric Vigoda. Improved bounds for sampling colorings. In *Foundations of Computer Science, 1999. 40th Annual Symposium on*, pages 51–59. IEEE, 1999.
- [18] Dror Weitz. *Mixing in time and space for discrete spin systems*. PhD thesis, Univeresity of California, 2004.